# Alternative WAIS to Search Online

by Ken Ficara

A professor announces an open-book exam and students show up with arm loads of textbooks and reference works. They soon discover, however, that the exam won't be the ace they bargained for. Because, unless they know exactly where specific information is located in each book, they won't have time to flip through all the sources in the hour or so allotted for the exam.

An online user can face the same situation. There are so many sources of information out there—each of which has to be searched separately—that half the battle is figuring out where to look. Is this information available internally, perhaps from your own corporate library? Or do you need to go to a commercial online service? If so, do you know exactly what each service has and does not have? Or will you need to check several?

Of course, you will have to know the different set of commands for each of the services you decide to check. Not only will you end up entering your request in different ways several times over, but you'll also have to review your results one set at a time.

**But what if** you could enter a search statement just once and get back relevant information from all sources—everything from your internal e-mail system to online services to publicly available free services—that had what you needed?

This is the idea behind Wide Area Information Server technology. WAIS (pronounced "ways") lets you search as many sources as you want without logging on to each one individually.

You select from a list of sources then enter your request once, in plain English. The WAIS server queries the selected sources, organizes the responses it gets, then presents them to you in one list. You review them, choose the ones that are the most relevant, and then send the server back to look for "more like these."

This means you could go online to a WAIS server located at supercomputer manufacturer Thinking Machines' headquarters in Massachusetts, and, through the server, simultaneously search nearly 400 sources—everything from MIT's collection of Aesop's Fables to a gathering of Supreme Court decisions at Oregon State University to your own word-processing files.

**The most important aspect of the WAIS** method of searching is that it does all the work: It accesses the different sources, enters the correct commands and codes, and organizes the results. All you do is choose articles that meet your requirements—and the server then retrieves them.

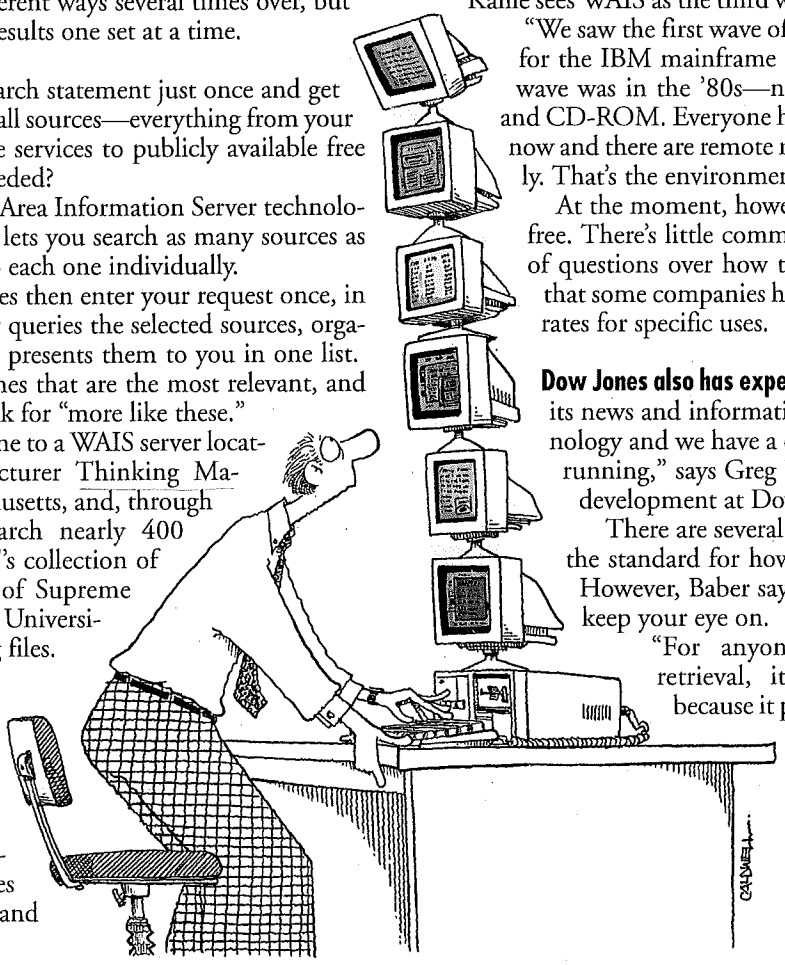
The search engine works much the way Dow Jones News/Retrieval's //DOWQUEST service does. //DOWQUEST, the first commercial service to use relevance feedback, lets you enter your search statement using plain English, then gives you a list of articles that fit the topic.

Once you choose the articles that are most relevant, //DOWQUEST goes off and finds ones similar to those you selected.

This is not experimental technology. WAIS Inc. of California sets up WAIS servers and helps people use the system. Brewster Kahle, president of WAIS Inc., says WAIS servers have 20,000 users in 28 countries, with 380 servers in 12 countries currently accessible. He says the number of users and sources doubles about every seven months and that the company's public-access WAIS systems have received more than 100,000 requests from more than 6,000 computer users world-wide.

Kahle sees WAIS as the third wave in online communications. "We saw the first wave of 'host-centric' systems designed for the IBM mainframe and dumb terminals. The next wave was in the '80s—networking, personal computers and CD-ROM. Everyone has PCs and Macs on their desks now and there are remote resources you can get to smoothly. That's the environment WAIS is designed for."

At the moment, however, all the available sources are free. There's little commercial use of WAIS yet because of questions over how to charge for it. But Kahle says that some companies have licensed their data at flat-fee rates for specific uses.



**Dow Jones also has experimented** with a WAIS server for its news and information. "We keep up with the technology and we have a couple of internal WAIS servers running," says Greg Baber, manager of research and development at Dow Jones.

There are several approaches that may emerge as the standard for how you get and use information. However, Baber says WAIS is an important one to keep your eye on.

"For anyone interested in information retrieval, it's a trend worth watching" because it provides a standard for searching that "will be important in the long-term goal of information at your fingertips," he says. //

*Ken Ficara is a database development associate in Dow Jones' editorial department.*

***For reprints of up to 20 articles, call Dowline at 609-520-4657.***
*Page numbers in blue signify a full profile of that database.*

12A-760
DOWLINE
PRINCETON, NJ
CIRC-90,000
MAGAZINE-QUARTERLY

3RD QUARTER 1993

Bacon's

10A-460
INFORMATION RETRIEVAL &
LIBRARY AUTOMATION

MONTHLY
NOVEMBER 1991

Bacon's

3481

# U.S. CONGRESS SERIOUS ABOUT STANDARDS, NISO CONFERENCE HEARS

The U.S. Congress has developed a new interest in standards development, according to Linda Garcia of the Office of Technology Assessment (OTA). Ms. Garcia is project director of an OTA study of U.S. involvement in standards development internationally and the impact of standards on U.S. competitiveness. Speaking at the annual conference of the National Information and Standards Organization (NISO), she said that this tenth study of standards is the most comprehensive that has ever been undertaken on international standards and their role in the marketplace. Two areas of focus are providing incentives to standards development and providing U.S. expertise to standards programs in developing countries.

Ms. Garcia said, "The Congress believes that it is time to address our policy making with respect to standards in order to understand the relationship between standards and the structure of the coming Information Age."

**Opening Up the Process**

The leader in the Information Age, according to Professor Michael Spring of the University of Pittsburgh, will be the standards organization that opens the standards-making process.

"We can no longer afford to write standards that may be ignored by the marketplace," he told the NISO conference. "Unless those who should apply standards have access to them, including in their formative stages, the standards adopted will not be used."

He offered three concrete suggestions for opening up the process: using electronics for meetings, draft reviews, and balloting; greater participation at the senior management level, particularly among managers in strategic planning and product development; and lowering "exorbitant prices" often charged for standards.

**The Discomfort of the Old Guard**

The politics of standards making is changing as the standards-making environment evolves, according to Dr. Rick Weingarten, executive director of the Computing Research Association. Pioneers in the field, he said, can be taken aback when there are new voices that insist on being heard--often without being very respectful of those who have gone before them.

"The old guard is uncomfortable," he said. "Instead of reaching out to the challenge, all too often there is a futile inclination to close ranks and try to maintain the status quo."

Dr. Weingarten urged the participants to be "flexible and open to new input."

**Access-Tool Give Away**

Demonstrating its openness and its commitment to one NISO standard, Z39.50, The Thinking Machines Corporation told the conference that it will give away its WAIS protocol specifications and source code to all interested parties. The supercomputer manufacturer is a key developer of the WAIStation (Wide Area Information Service) Project, an experiment automating the search and retrieval of many types of electronic information over wide area networks.

So far users may only use the WAIS protocol to access data banks in the public domain, though Thinking Machines hopes to change this.

"We believe that standardization is the critical link to the next plateau in the Age of Information," project manager Franklin Davis said, "and we hope that giving WAIS away will demonstrate this and catalyze a market for information servers."

## VIDEO SYSTEM PLAYS 32 TAPES AT A TIME

Visitors to the Museum of Television and Radio have access to the museum's 70,000 tape collection, 96 viewers at a time. The museum, which opened in New York City in September, is using three units of a videotape storage and retrieval system that plays back video programs for up to 32 users at one time.

The VLCS-800 from Sony comprises three parts: a multicassette playback console, a control system, and user terminals. A robotic device automatically locates the cassettes containing the desired program. Once the cassette is loaded into the VCR, the system advances the tape until it senses video. It then switches audio, video, and VCR control to the user. The total system occupies the floor space of a standard office desk.

The VLCS-800 uses the video Hi8 format. This format differs from the standard 8 mm format with signal improvements that provide more than 400 television lines of resolution. Program selection and VCR remote control are prompted by the viewing monitor and performed using a keypad.

The system provides a variety of features necessary for cassette management, including usage and error

UMAX Technologies
Clara, CA 95051; ph
0771; fax, 408-982-0776

## APPLE & IBM CONCLUDE AGREEMENTS

Apple and IBM have released the final details of their recent alliance. It will consist of five distinct technology initiatives: three to expand the companies' current technologies and two to focus on "the creation of new foundation technologies."

The current technologies to be expanded are Macintosh and IBM systems networking; RISC (Reduced Instruction Set Computing) microprocessors; and open-systems platforms. The first networking products are expected as early as December. Together with Motorola, Apple and IBM will create a new family of RISC microprocessors for personal computers and workstations to be available in 2 to 3 years. PowerOpen will be a new open-systems environment made possible through licensing agreements and also to be available in 2 to 3 years.

The two new technologies to be jointly explored are multimedia and object-oriented software. An independent multimedia company will develop, license, and make available specifications and technologies to promote the exchange of information such as sound, graphics, video, text, and animation across a variety of computing and consumer electronic devices.

The independent object-oriented software company will develop and license "an entirely new genre of system software that greatly simplifies computer programming and allows much easier customization of software programs," according to Apple. The company's technology will be used by both Apple and IBM as the core of new products for the 90s.

## FAXON RECEIVES LIVE TRANSFER OF STANDARD DISPATCH DATA

The Faxon Company has announced that it has received and uploaded the first transmission of live standard dispatch data in the serials community. Kluwer Academic Publishers in Dordrecht, The Netherlands, produced the data stream and Faxon received it as participants in a X12 pilot project of the International Committee for EDI in Serials (ICEDIS).

To effect the transmission, the dispatch data are coded in the ANSI X12 standard by Kluwer using the

# Focus On WAIS

**Continued from page 44**

protocol stack.

As shown in Figure 2, I decided to search my mail archives and technology articles for references to NCD's Xremote protocol. The results appear in the scrolling list. If the result is an entire file, such as the article contained in the file "XSerialArticle.txt," the path name for the file is listed after it.

The other results in the list are individual E-mail messages that actually are in several large text files on the WAIS server. Because the WAIS indexing program understands E-mail format, it was able to index individual E-mail messages in my E-mail archive files and transfer only those E-mail messages pertinent to the client query.

By clicking on a document in the Results window, the portion of the result most relevant to my query appears in another window. The *waisserver* uses a simplistic approach to interpreting my request for information about Xremote. It looks for the word "xremote"—the search is case-insensitive—in mail messages and headers and displays matching documents and mail messages in the results window. This turns out to be adequate as long as you put meaningful words in your query.

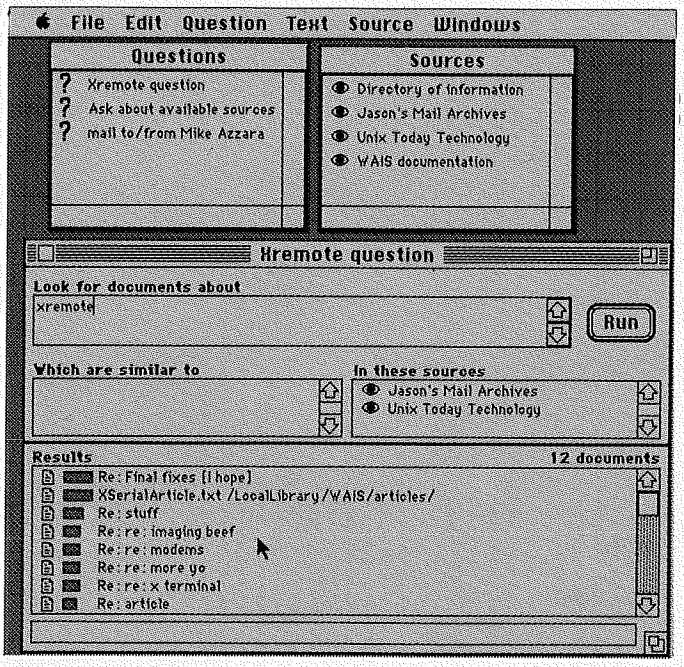



**Figure 2: Mac WAIS client shown with results of a search for "xremote"**

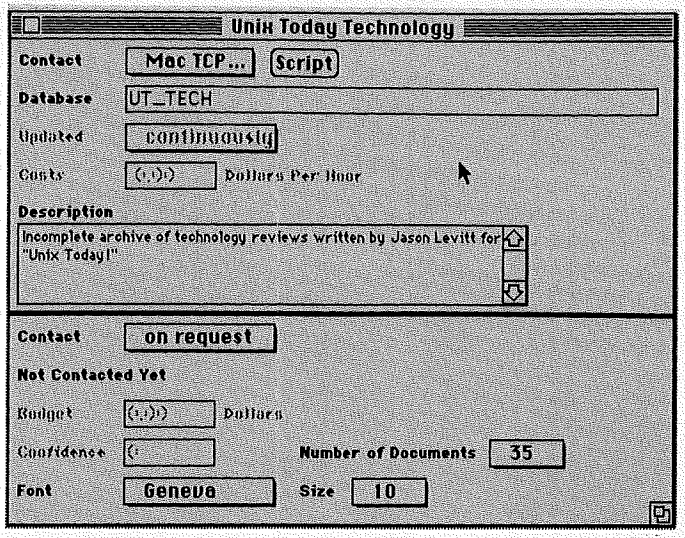



**Figure 3: WAIS client set up to use Mac TCP/IP and the database UT_TECH**

TMC has a much more sophisticated searching mechanism in its Internet server, *quake.think.com*; however, the search source code is not freely available.

One of the key features of the WAIS protocol is its ability to allow secondary search criteria. In Figure 2, the criteria would be entered by copying a result, or chunk of a result, to the "which are similar to" window. A subsequent search would use any words contained in that window as additional search criteria. Repeatedly using that method can quickly refine the search parameters.

## AN OPEN END

The next version of the WAIS protocol should be officially folded into the Z39.50 standard this month and is expected to include multimedia support and integral support for English-language queries. These enhancements should add considerable clout to WAIS, given the infant state of commercial multimedia query/retrieval technology.

WAIS software is freely available from a number of sites. Unfortunately, the WAIS client program can only be obtained via anonymous FTP at this time, which means you have to have direct Internet access.

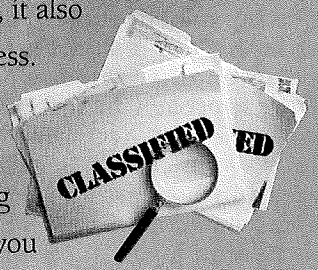
The WAIS server and X-based client program for Unix are available on *uunet.uu.net* in the directory */networking/distrib-is/wais*.

My small network experiment with WAIS only touched on its full potential; however, for my small database needs, it was quite useful. The free WAIS software is, like the MIT X software, meant as refer-

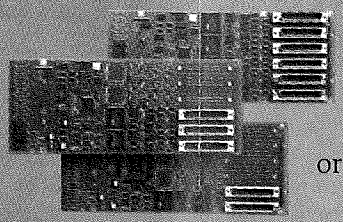
**Continued on page 48**

# Money, security, and the right connections: everything you'd expect from a distinguished family.

Annex™ communications servers provide a combination of price, data security and performance that no one else can match. The Annex Three, for example, delivers 38.4 kbps on all ports. At less than $110 per port in a 64-port configuration, it also delivers the lowest price per port in the business.




If your connectivity needs are more decentralized, consider the Micro Annex XL. Available in 8- or 16-port configurations, it provides the same advanced networking features as the Annex Three at a lower cost per server. If you demand Annex quality on a strict budget, the Micro Annex ELS gives you the functionality you need at the lowest possible price.



To find out more about the world's most distinguished family of communications servers, contact your authorized Xylogics reseller today or call us directly at 800-225-3317.

**Xylogics®**

Xylogics U.S.A., 53 Third Avenue
Burlington, MA 01803
(800) 225-3317

Xylogics U.S.A., 18 Park Avenue
Los Gatos, CA 95030
(408) 354-4122

Xylogics International,
Featherstone Road, Wolverton Mill,
Milton Keynes MK12 5RD, U.K.
+44 908 222112





**Circle Reader Service No. 48**

# Spotlight On WAIS

**Continued from page 47**

ence software for further development, not as a commercial-quality implementation.

I encountered bugs, such as a persistent permissions error from the NeXT client, and strange window clipping from the Motif client, and I have yet to get the *waisserver* running cleanly under SVR4. But when the software is open and free, who cares?

The vision of WAIS is not only easy access, retrieval and publishing of information, but the creation of a marketplace that can encourage new information sources.

That, according to advocacy groups such as the Electronic Frontier Foundation, could be realized through ISDN, an infrastructure for a "National Public Network" that already is partially implemented in the U.S. telephone system. Such a network could bring the reality of WAIS-based online information services into virtually every home.

tact any of the Authorized Distributors listed below.

**Almac/Arrow Electronics**
**14360 S.E. Eastgate Way**
**Bellevue, WA 98007**
**800-426-1410**

**Avnet Computer**
**10950 W. Washington Blvd.**
**Culver City, CA 90230**
**800-426-7999**

**MTI Systems, Arrow Electronic**
**25 Hub Drive**
**Melville, NY 11747**
**800-645-6530**

**Pioneer Standard**
**4800 131st Street**
**Cleveland, OH 44105**
**800-874-6633**

**Pioneer Technologies**
**15810 Gaither Road**
**Gaithersburg, MD 20877**
**800-227-1693**

**Wyle Laboratories**
**128 Maryland Street**
**El Segundo, CA 90245**
**800-289-9953**

# The Promise Of The WAIS Protocol

## *Emerging Standard Represents First Step Toward Unifying Data Search & Retrieval*

**BY JASON LEVITT**

It doesn't take an expert to see that the state of modern information handling is neither open nor unified. A trip to the main library at the University of Texas at Austin—one of the top 10 college library systems in the U.S.—confirms this.

The primary card catalog is contained on an IBM mainframe accessible through various synchronous block-mode terminals scattered about the main library, and also accessible via modem through a rather crude dial-up facility.

In the main reference room, an OCLC (On-line Computer Library Center) terminal allows access to other university card catalogs; several IBM PCs are available to search CD-ROMs for bibliographical citations and abstracts on a variety of subjects; and a LEXIS/NEXUS terminal can be used for researching major U.S. court decisions. In the engineering library, an IBM PC with CD-ROM is available for searching U.S. patents.

### WAIS Server Source File

```
( :source
  :version 3
  :ip-name "nextbox.utoday.com"
  :tcp-port 5001
  :database-name "UT_TECH"
  :cost 0.00
  :cost-unit :free
  :maintainer "jason@nextbox.utoday.com"
  :description "Server created with WAIS release
8 b2 on Mon Nov 18 16:54:19 1991 by
jason@nextbox.utoday.com
UNIX Today! technology articles by Jason Levitt
The files of type text used in the index were:
  /LocalLibrary/WAIS/articles/ABCstory.txt
  /LocalLibrary/WAIS/articles/AIX3.1FS.txt
  /LocalLibrary/WAIS/articles/Benchinfo.txt
  /LocalLibrary/WAIS/articles/LPFstory.txt
  /LocalLibrary/WAIS/articles/MacXstory.txt
  /LocalLibrary/WAIS/articles/Solbourne.txt
  /LocalLibrary/WAIS/articles/SunStory.txt
  /LocalLibrary/WAIS/articles/XSerialArticle.txt
  /LocalLibrary/WAIS/articles/Xarticle.txt
  /LocalLibrary/WAIS/articles/Xcontrib.txt
"
)
```

Figure 1

If one were to compare information accessibility at this facility to computer resource accessibility, things here are still in the early 1980s or late '70s. Each of the systems mentioned are primarily standalone and proprietary, having their own information retrieval and organizational formats with little, if any, interoperability between databases.

While the monster mainframe card catalog might provide pointers to many sources, it is ignorant of most other on-line sources and almost never provides the most current information on subjects, despite the best efforts of its administrators. What these information-handling systems need is a dose of open systems standards and technology, the same technology that is changing the face of modern computing.

Enter WAIS, for Wide-Area Information Server, a fledgling step in the overwhelming effort needed to unify information search and retrieval technology. WAIS is an emerging open systems standard protocol for query and retrieval of information. WAIS, pronounced "ways," is the brainchild of Brewster Kahle, an employee of Thinking Machines Corp. (TMC), the No. 2 supercomputer manufacturer, behind Cray, and purveyor of fine, massively parallel systems.

The basis for WAIS is the rapidly growing electronic-publishing movement, which is seeing more and more materials, usually available only in book form, "published" or placed onto electronic media such as disk and tape, where it can be accessed with a computer.

### WAIS TECHNOLOGY

WAIS is a protocol for the transmission of query and retrieval information, much like the information you would use to search a library card catalog. It is, in fact, an extension to an existing protocol standard called Z39.50, the Information Retrieval Service Definitions and Protocol Specification for Library Applications.

The Z39.50 standard was created by a group called NISO, the National Information Standards Organization, and is designed for use in electronic library card catalogs. Z39.50 essentially specifies formats for search requests directed at a database and formats for document retrieval requests. WAIS extends the Z39.50 standard to allow, among other things, discrete portions of documents, called "chunks," to be retrieved. This is especially useful in low-bandwidth situations such as serial links, where transferring an entire document in response to a query would be prohibitively time-consuming.

The WAIS protocol fits neatly at the top of the ISO 7-layer protocol model at the application and presentation layers. This makes it extremely portable to differing network environments such as TCP/IP and X.25.

★ Text ★ Retrieval

Like any good open standard, the WAIS protocol does not specify or limit the technology at either end of the wire. A WAIS client can be as simple as a command line interface that takes a database name, network address and query string as input, or as complex as a combination spreadsheet and database that constantly updates in real time, based on client/server activity taking place in the background. The only condition is that the client and server exchange query and retrieval information using the WAIS protocol.

The free WAIS source code, discussed later, implements a very typical client/server model for Unix-based Internet applications. The server creates and waits on a socket attached to a well-known port. Clients attach to the port using the port number and network address of the machine. The server accepts a request, forks a child process to handle the request, and then continues to wait and service other requests.

Requests for information are largely governed by special text files maintained by the WAIS server, called "sources," that vaguely resemble library catalog cards. Figure 1 shows a source I created containing 10 of my previous technology articles for *UNIX Today!* There is enough information in the source structure, network address, TCP port number and database name for any other machine on the network running a WAIS client to locate, understand and access the information in the database.

Not surprisingly, WAIS is already being used to connect archive sites on the Internet running on various Unix-based machines as well as proprietary systems such as Macintosh and NeXT. According to Brewster Kahle, there are approximately 80 sites running public WAIS servers and many more running WAIS privately within corporations and academia. A FidoNet WAIS server site was recently added to this collection of public sites running SLIP over a 9,600-bps serial link.

### FREE WAIS SOFTWARE

I like software that you can use to get some meaningful work done quickly without having to dig too deeply into documentation. The freely available WAIS software fits that description. In the *UNIX Today!* labs, I decided to put together a small heterogeneous network and run WAIS.

Acting as the WAIS server system (and also a client) was a NeXTstation. Attached over Ethernet was a Macintosh running MacOS and a Sun 3/60 running SunOS 4.1. The free WAIS software included NeXT and Mac binaries and complete source code for the Unix systems, in this case the Sun. I dug out my archives of personal Unix electronic mail, about 10 Mbytes' worth, and used the indexing program included with the WAIS server to create a hashed database. I did the same with 10 of my old technology articles written for *UNIX Today!* The databases, or "sources," are listed in Figure 2.

The WAIS indexing program knows about the format of many common types of structured online data such as electronic mail, *netnews*, PICT-/GIF/TIFF files and biology abstract formats, and it also handles straight ASCII text.

There was also a database of WAIS documentation, created automatically by the server program, and a directory of all sources I created called "directory of information" that simply points to all the databases. After creating the databases, I ran the WAIS server program, called *waisserver*, on the NeXTstation, which sits and waits for incoming WAIS client requests.

Once the *waisserver* was running, I could access it using the clients, called WAISstations. On the Sun, which was running X/Motif, I chose to use the Motif client. I also used the Mac and NeXT WAISstations. In order to access a *waisserver*, I first had to set up my sources. Figure 3 shows a source setup window for the Mac client. I had named my database of articles "UT-TECH" on the WAIS server. The access method, "Contact," was MacTCP, Apple's TCP/IP

Continued on page 47

### General WAIS Information

**Thinking Machines Corp.**
1010 El Camino Real, Ste. 310
Menlo Park, CA 94025
415-329-9300 Fax: 415-329-9329

Bibliography of available WAIS documents.
Send electronic mail to: barbara@think.com

**Accessing a WAIS client on the Internet**

Telnet to quake.think.com, login as **wais**

**Getting involved with the Nat'l Public Network**

**Electronic Frontier Foundation**
155 Second Street
Cambridge, MA 02141
617-864-0665
E-mail: eff@eff.org

### On-Line WAIS Discussions And Development

alt.wais newsgroup on USENET

Join mailing lists by sending e-mail to:

wais-discussion-request@think.com – Weekly digest of mail from users and developers

wais-interest-request@think.com – Infrequent announcements of new releases

wais-talk-request@think.com – Developers' mailing list

**Free WAIS client software**

Clients for NeXT, X, Macintosh, Unix ASCII, GNU Emacs and Motif.
Anonymous FTP to think.com in the directory /wais

Clients for VMS, MS-DOS, Novell LAN Workplace and SunView.
Anonymous FTP to samba.oit.unc.edu in the directory /pub/wais/UNC

**Free WAIS server software**

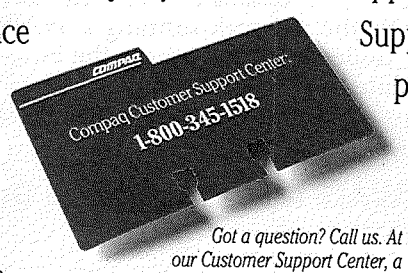
Servers for NeXT and various Unix platforms
Anonymous FTP to think.com in the directory /wais

# SYSTEMPRO/LT Family. Priced for small budgets.


Setup is easier, too. COMPAQ Integration ToolKits tell you just how to do it. We've already tested and proven reference platforms—combinations of hardware and software that you'd likely use—including Banyan, Microsoft, Novell and SCO UNIX operating environments. We know the ropes.

Once you're up and running, the new COMPAQ System Manager keeps you in control. It continually monitors server performance and instantly notifies you if something is wrong. You can even diagnose and correct problems from remote locations.

If you have a problem with one of our reference platforms, we'll get it solved. No buck-passing. Through Technical Support Alliances, we and leading hardware and software developers cross-train our support people so we can help you better. Our Technical Support Line gives subscribers extra help with complex problems. And spare parts can be shipped to your dealer within the same day.




*Got a question? Call us. At our Customer Support Center, a bank of knowledgeable people is ready to help.*

Learn more about how the COMPAQ SYSTEMPRO/LT Family can serve your workgroup's needs and budget. Call for our free brochure and the location of an Authorized COMPAQ Computer Dealer with expertise in your preferred operating environment. Just call 1-800-231-0900, Operator 177. In Canada, you can call 1-800-263-5868, Operator 177.

COMPAQ

It simply works better.

3481

# Beyond the Theory: Knowledge Networking in Practice

*by Todd Miller*

Unless you have been living in a cave with an unlisted b-board mailbox, you are well aware of the current, hottest, trend in the information industry, that toward database access via OPAC systems, or Knowledge Networking. A couple of years ago, industry cognoscenti foretold the coming of the network and the impending arrival of its long-awaited by-product, the "library without walls."

While indeed the walls are tumbling down across the country, some of the homeless are no doubt wanting for shelter from the numerous "little details" which torment the installation of these networks. These are the items typically dispatched with waves of hands when decision makers and sales representatives meet.

Fortunately for the many libraries which have not yet implemented Knowledge Networks, enough real experience has been accumulated in recent months to provide some clues on what to think about when planning to install a network. This will allow many to benefit from the occasional misfortune of the brave infopioneers.

## Strategic Issues vs. Anal

There's no question that it is more fun to waft in images of imaging than it is to wallow in disk drives. Nevertheless, the seemingly mundane issues determine the utility of a system in the 20th century, while establishing a solid foundation for the 21st. An ideal installation will sweat the details of the present, yet have sufficient awareness of the general direction in which the future is headed in order to grow in an intelligent way. Recent experience has revealed a number of obvious landmarks and hazards. While it is unlikely that a couple years of experience can expose all possible obstacles, if those listed here receive careful attention, the resulting system will be well equipped to take on the unknown.

## Disk Space

Quite simply, there's never enough. On the surface, this may seem puzzling, as this ought to be one of the more easily quantifiable items on the Network shopping list. The problem lies primarily in indexing overhead. In the process of making the database navigable by the machine (as well as the user), the data must be "indexed," a term not used with true accuracy in the library sense. The process involves establishing logical links between selected terms in the database, and their destination information records. The greater the number of terms used, the greater the overhead, and the more disk space consumed. Some automation vendors' indexing schemes will grow a database as much as ten times, turning a Davidian database like Magazine Index of 400 megabytes into a Goliath 4 Gigabytes!

One must question the utility of a database which has been grown this large by over-indexing. While it is arguable that subject, author, and title indexing of these fields in the bibliographic citation is important, it is probably also true that linking the CODEN field to a full text is going a bit too far.

## To Print, or Not to Print?

That is the question for the vast majority of public-access terminals, which at present have no companion printing devices. Now that we have the ability for our patrons to search and display a state-of-the-art information service, do we utilize number 2 pencils as the output device?

Then there is the question of full text and images. While the idea of dedicating laser printers to hundreds of public access catalog workstations is enviable, the current cost of the technology makes it less so. Printer servers may make the most sense over the long run. The thrift plan may entail small groups (2-4 on a single table) of terminals tied to a dot-matrix printer with an automatic switch, with larger groups of terminals feeding to laser printer centers. Of course, one must then contend with sorting articles from multiple patrons and locations.

## Expertise

Academic libraries represent the lion's share of those which have installed Knowledge Networks to date. The reason for this lies primarily in the ready expertise available from computer science departments at many major (and not-so-major) universities. While the marketing literature may paint a lovely picture of turnkey networking, the truth is that the state of the art is unpolished, and some assembly is required.

## Vendor Support

There are two varieties here, information vendor and automation vendor. The information vendor should be able to provide the following: data in MARC format, authority file/subject guide in MARC format, and documentation, which fully describes the size of the file in terms of bytes, number of records, and average size per record. Documentation should also completely describe the field specifications of the database and search mechanism.

Ideally, the automation vendor should provide either a pre-indexed database, or reasonably-priced network access to the data. A small handful of vendors offer the former service, and at this time, only two offer a network service to their customers, CARL and Data Research Associates. Others are known to be considering providing such a service. The major-

*(continued on next page)*

18A-1920
INFORMATION TODAY
10.000

MONTHLY
MAY 1992

WAIS

# Beyond the Theory

*(continued from previous page)*

ity of vendors ship software to their customers, and the customer performs the processing necessary to load and index the database. A number of library networks obviate the need for vendor support by subscribing to the data and providing their own software. The University of California MELVYL system and the Florida Center for Library Automation are examples of this approach. These systems load data in one central location, then broadcast it to their member libraries, eliminating the need for vendor involvement. Obviously, this approach requires a relatively high level of sophistication on the part of the library system. and is not recommended for the general population.

## Start Your Search Engines

The utility of the database searched on the network varies directly with the effectiveness of the search engine and its user interface. Most vendors offer a keyword search capability, with the ability to restrict searches using Boolean operators. The most common fields searched are subject, author, and title. The majority of automated systems at present do not allow the ability to search by way of a cross-referenced subject guide or authority file, though several are working on this. This is especially important for periodical indexes which are indexed by humans before undergoing the OPAC's "indexing" process. Ideally, a system should use one authority for all materials searched, whether they be books, periodicals, encyclopedic material, directories, etcetera. The user probably doesn't really care about the source of the information, as long as answers are provided to his/her questions.

Related to the efficacy of the software is the user interface. Since most terminals utilize an ASCII display, the options for the presentation of the interface are few. The advent, however, of bit-mapped displays will not only provide for the presentation of images, but for creative Graphical User Interfaces (GUIs—pronounced "gooeys") as well. Ideally, these will overcome much of the computer anxiety experienced by novice users.

## WAIS and Means

Once the subject-access basics of searching are squared away, it is well worth understanding the growth path of your system's search engine and interface in order to avoid obsolescence. Many experts believe that the engines of the not-so-distant future will be asked to discreetly inquire beyond the boundaries of their immediate library's system in the course of satisfying patron data demands. The discretion will come in the form of Z39.50, an applications-layer protocol within the Open Systems Interconnection (OSI) reference model[1]. Put very simply, Z39.50 allows different machines with different operating systems to transparently exchange data. Imagine searching a locally-mounted periodical index on your librarys' DEC VAX. The index record indicates that text is available for this record. You press a button et voila!, the text appears. Only this text was dispatched from an IBM 2,000 miles away! This is a simple example of what Z39.50 will do. A handful of automation vendors are currently at work building Z39.50 engines, and it is likely that any system which does not speak fluent Z39.50 by the mid-'90s will be considered inadequate and obsolete.

## Dig Those Heuristical Algorithms

It will also enable virtual search environments like The Wide Area Information Server (WAIS—pronounced "ways") to become reality[2] WAIS, an experimental project of Thinking Machines Corporation in Cambridge, Massachusetts, will search any compatible system on a network, then prioritize the results, based on simple heuristical algorithms. Prioritizing search results according to degree of relevance will become increasingly important over time as the growing number of compatible nodes on a virtual network dump more and more data into the ocean of navigable information. The point is that the system must be scalable, allowing for the growth of both the complexity and number of clients and servers.

## Can We Talk?

Of course, all this necessitates that your system network (and that your network work!). Hopefully by the time Z39.50 interfaces have pervaded OPAC systems, these machines will have the physical means to communicate with one another. It is likely that future communications will happen via NREN, the National Research and Education Network born of the High Performance Computing Act of 1991[3]. NREN was first proposed by Senator Albert Gore, Jr. of Tennessee (Senate Bill 272) passed on September 11, 1991. NREN is destined to replace the venerable Internet, originally established by the National Science Foundation in the late '80s[4]. While a specific definition has not yet been established for NREN, it will probably use the SONET-based Advanced Network and Services, Inc. (ANS) as the backbone for the network[5]. ANS was founded last year

*(continued on next page)*

# Beyond the Theory

*(continued from previous page)*

by IBM, MCI, and Merit, Inc. While this system will probably provide adequate bandwidth for the transmission of images, it falls short of the 600-1,600 M-bps asynchronous transfer rate called for by the initiators of NREN. Regardless, NREN still represents a tremendous improvement in machine-to-machine communications relative to Internet, whose 13 hub backbone is limited to throughput of 1.5 M-bps[6]. Bidding is slated to commence this month to upgrade the backbone to a 45 M-bps fiber-optic network, however[7]. The point of all this is that while it may not be essential that your OPAC engage in mixed-vendor relationships today, this time is drawing near. For this reason, it is important to be aware of your vendor's plans for communications, particularly as they apply to Z39.50.

## Hierarchy of Demand/Hierarchy of Supply

The longer term technical issues dealing with networking and virtual search capability demand attention to larger information management concerns. The growing number of sources of electronic information makes it virtually impossible for a single library to house all available data. This has two implications for the library: 1) the library must decide which information will reside at the library and which will not, and 2) the library system must be capable of going beyond its immediate system and onto networks for information considered esoteric to, or simply in low demand by, its patrons.

Deciding which information should live at the library involves looking at the information demands of the library's constituency, and understanding what information must be delivered instantaneously vs. that which can be delivered in a minute, an hour, or day. This shopping list must then be compared with the relative costs of storing the data locally, vs. accessing it through a network, interlibrary loan, or traditional online services. One tool which can assist this process is existing subject and source usage information obtainable from like libraries with well-instrumented systems.

> *"An ideal installation will sweat the details of the present, yet have sufficient awareness of the general direction in which the future is headed in order to grow in an intelligent way."*

While this presents a complicated decision matrix for the information professional, this may not be as difficult an issue as that of giving up ownership to the data. There appears to be a prevailing sense that this is one of the library's fundamental "value addeds." Certainly, when information sources were more limited then they are today, this was true. The issue today, however, is not one of capturing all information available from a severely limited universe. Rather, the issue appears to be one of sorting through an *excess* of data to find that which is most relevant to the patron. It represents a fundamental evolution in the role of the library, from information warehouse to information navigator.

This makes construction of the network all the more challenging. It would be difficult enough to attend to the anal issues of disk drives and printers, without being forced to plan a comprehensive information and technology matrix while the world is simultaneously turning upside down. To make matters worse, all this requires a considerable commitment in technology and dollars to an information future, which is at once vague and untraditional.

It might be nice if the world were a static place, and one had infinite time to sort through extremely complex issues such as these. Unfortunately, opportunity windows are dynamic, and patrons demand basic services even while oracles are being constructed. Because every library is different, there is no one solution which can be mass produced. There are, however, case experiences, conventional wisdom, statistics, and opinions on what is needed and where things are headed—valuable data to support what, in the end, still boils down to a gut feel decision on what is best for your library.

---

*Todd Miller is Senior Manager of Library Automation at Information Access Company, where he manages the company's database site licensing program.*

**Footnotes**

1. Clifford A. Lynch. "Information Retrieval Service Definition and Protocol Specifications for Library Automation."
2. Brewster Kahle, Art Medlar. "An Information System for Corporate Users: Wide Area Information Servers," *Online, September 1991.*
3. Scott Mace. "Congress Gives Nod to NREN," *InfoWorld, September 23, 1991.*
4. Sharon Fisher. "Whither NREN?" *Byte, July 1991.*
5. Thomas J. Hargadon. "NREN, a Fundamental Turning Point in Networks?" *The Office, July 1991.*
6. David Churbuck. "NFS To Solicit Backbone Operators," *Computerworld, December 9, 1992.*
7. Gary Anthes. "NFS To Solicit Backbone Operators," *Computerworld, December 9, 1992.*

3481
INTEROPERABILITY

# WAIS: Is It the Lotus 1-2-3 of the Internet?

*San Francisco*—Before you can sell a computer, you need an application. Until you get that magic Lotus 1-2-3 or Aldus Pagemaker, the market won't explode. The same goes for networks.

A new application called WAIS may prove to be the kicker for the Internet. Mind you, the Internet hasn't really needed a kicker. It doubles in size every seven months.

If you look at what people are doing on the Internet, though, you see the same old, same old. Electronic mail, FTP for file transfer, and Telnet virtual terminals are the three fundamental applications. The computing paradigms are basically the same ones invented in the early days of the Arpanet in the 1970s.

Carl Malamud

Brewster Kahle of Thinking Machines Inc., Cambridge, Mass., led a team of people that has developed a totally different way of using the network, a group of cooperating applications called the Wide Area Information Servers.

WAIS (pronounced like "weighs") is a way of finding and retrieving documents on the Internet. A document can be simple ASCII text, or can incorporate graphic images or be generated dynamically as a result of a database query. The server accepts queries using the Z39.50 protocol and returns names of documents and the content of those documents. A typical server accepts queries in natural English, translating the query into key words with which to search the database.

Clients can talk to lots of different servers on the Internet. A single query may be sent out to many places, and documents from different sources collated, allowing the user to view the network as one big database.

What makes this platform interesting is the kind of servers and client interfaces that have developed. Servers now available have an incredible range of information and are run by a gamut of groups. Interested in religion? You can dispatch queries to servers that have the full text of the Koran, the Book of Mormon, and the Bible.

Need information about water quality in Alaskan river basins? Servers ranging from the Columbia Law Library to a global weather change master directory will find dozens of relevant documents.

Client interfaces to WAIS have been developed for the Macintosh, X Windows, Next and several other graphical platforms. The client software lets a person compose queries, refine them, and then save the queries for periodic re-execution. You can think of this as a personalized newspaper.

WAIS can be run on a local network as a corporate information service. Or, for an easy entry to WAIS, a few workstations or PCs can access the servers already available on the Internet. Most effective is to do both, running a local server for corporate information and hooking up to the Internet to open up a broad universe of info for employees.

Why would Thinking Machines develop WAIS? One can think of Thinking Machines' massively parallel processors as the ultimate WAIS servers. Different processors can all look for different things, letting complex searches of large databases be run quickly. ■

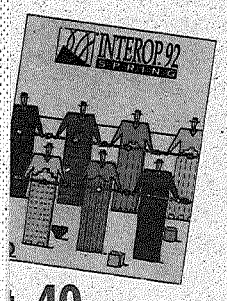
*CARL MALAMUD can be reached via the Internet at* carl@malamud.com. *The opinions expressed here are his own.*

**ORK**
**IG**
**ISDN?**

d to a function-
le opportunity for
ions.

**strations** - put it all
ame Relay, X.400,
l SMDS in action.

u're going, long
tend INTEROP
ngton, D.C. And
one.




**. 40**

US, FAX 1-415-949-1779